

# Deep Learning, Feature Learning, and Clustering Analysis for SEM Image Classification

Rossella Aversa<sup>1,2</sup>, Piero Coronica<sup>1,3</sup>, Cristiano De Nobili<sup>1,4</sup> & Stefano Cozzini<sup>1,5†</sup>

<sup>1</sup>National Research Council-Istituto Officina dei Materiali (CNR-IOM), 34136 Trieste, Italy

<sup>2</sup>KIT-Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

<sup>3</sup>Research Software Engineering, University of Cambridge, Cambridge CB3 0FA, UK

<sup>4</sup>Freelance at denocris.com

<sup>5</sup>Area Science Park, Padriciano 99, 34149 Trieste, Italy

**Keywords:** Neural networks; Feature learning; Clustering analysis; Scanning Electron Microscope (SEM); Image classification

Citation: R. Aversa, P. Coronica, C. De Nobili & S. Cozzini. Deep learning, feature learning, and clustering analysis for SEM image classification. *Data Intelligence* 2(2020), 513–528. doi: 10.1162/dint\_a\_00062

Received: September 6, 2019; Revised: July 27, 2020; Accepted: July 30, 2020

---

## ABSTRACT

In this paper, we report upon our recent work aimed at improving and adapting machine learning algorithms to automatically classify nanoscience images acquired by the Scanning Electron Microscope (SEM). This is done by coupling supervised and unsupervised learning approaches. We first investigate supervised learning on a ten-category data set of images and compare the performance of the different models in terms of training accuracy. Then, we reduce the dimensionality of the features through autoencoders to perform unsupervised learning on a subset of images in a selected range of scales (from 1  $\mu\text{m}$  to 2  $\mu\text{m}$ ). Finally, we compare different clustering methods to uncover intrinsic structures in the images.

---

## 1. INTRODUCTION

Image classification, as well as image recognition, image retrieval and other algorithms based on neural networks, are widely applied in many different research areas, including nanoscience and nanotechnology [1, 2]. Scientists working with microscopy techniques are particularly interested in general tools able to automatically identify and recognize specific features within images.

---

<sup>†</sup> Corresponding author: Stefano Cozzini (Email: stefano.cozzini@areasciencepark.it; ORCID: 0000-0001-6049-5242).

Neural networks were employed in a number of recent studies for feature extraction from different types of microscope images. For example, recognition of cellular organisms from scanning probe microscopy images was shown using artificial neural networks [3]. Neural network classifiers were also used to estimate the morphology of carbon nanotube structures, such as their curvature and alignment [4]. In the framework of microscopy cells images, [5] presented a method for cell counting based on a fully convolutional neural network (CNN) able to predict a spatial density map for target cells; a similar method was used by [6] for candidate region selection and for further discrimination between target cells and background; in [7], a supervised CNN was trained to identify spots in image patches.

In nanoscience, where large numbers of images are the typical outcome of experiments, image recognition techniques can be an extremely powerful tool. In the framework of the NFFA-EUROPE project [8], the information and data management repository platform (IDRP) was developed to suit the data sharing needs of the nanoscience community. The first aim of this distributed research infrastructure is to provide a common repository where NFFA-EUROPE scientists can easily collect and store scientific data produced at the experimental and theoretical facilities available among the project partner institutes. The second and even more important purpose is to allow the users to share and publish the collected data according to the FAIR principles [9]. A central problem is thus to guarantee the search and retrieval of such heterogeneous data, and to establish a proper way to organize the repository: this makes almost mandatory to provide tools which automatically enrich data with appropriate metadata defining its content.

We focus on the data produced by a single instrument, the Scanning Electron Microscope (SEM). This is an extremely versatile instrument, routinely used in nanoscience and nanotechnology to explore the structure of materials with spatial resolution down to 1 nm. Almost 150,000 images were collected in the last five years by the TASC laboratory at CNR-IOM in Trieste [10] and such number will increase in the near future. We thus face the problem to classify and store them in a FAIR way.

As a fundamental step, a sample counting more than 18,000 images was extracted from the original sample and manually labelled in 10 categories, forming the SEM data set [11], which we employed in [12] and in [13] to investigate transfer learning [14], in particular feature extraction, for automatic image categorization. The test accuracy resulting by the adopted technique settled around 90%.

In this work we aim at improving the accuracy in the classification task through an extensive comparison among three well established CNNs and different machine learning techniques [15]. As a further scientific development, we also face the challenge of improving the existing categories by means of a semi-supervised approach to automatically find hidden structures in the data [16]. This is the first step towards the automatic addition of new categories and the creation of a hierarchical tree of sub-categories, reducing the huge human effort required to manually label the training set.

The paper is organized as follow: in Section 2 we show the supervised approach using CNN networks. We also introduce a further improvement which allows classifying SEM images in terms of their scale. Section 3 presents our unsupervised approach to the problem, discussing why a completely unsupervised approach for feature learning is not possible in this case. Finally, in Section 4 some conclusions and future perspectives are presented.

## 2. SUPERVISED LEARNING

The goal of this Section is to illustrate the techniques we adopted to increase the accuracy of the image classifier presented in [13]. We first trained from scratch different state-of-the-art CNN architectures on the SEM data set, and then we went further applying the following transfer learning methods:

- **Feature Extraction:** Start with a pre-trained checkpoint, reset and randomly initialize only the parameters of the last layer. Then, retrain the network allowing back-propagation just on the last layer;
- **(Complete) Fine Tuning:** Start with a pre-trained checkpoint, reset and randomly initialize only the parameters of the last layer. Then, retrain the network allowing back-propagation through all the layers.

We adopted checkpoints pre-trained on two data sets: ImageNet [17], a large visual database designed for object recognition, in the version of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012 [18], and our own SEM data set [11]. We note that the second case cannot be formally defined as a *transfer learning* technique, since the fine tuning of the CNN is performed on the same data set of the checkpoint; nevertheless, this is a commonly adopted way to efficiently refine the parameters of the network.

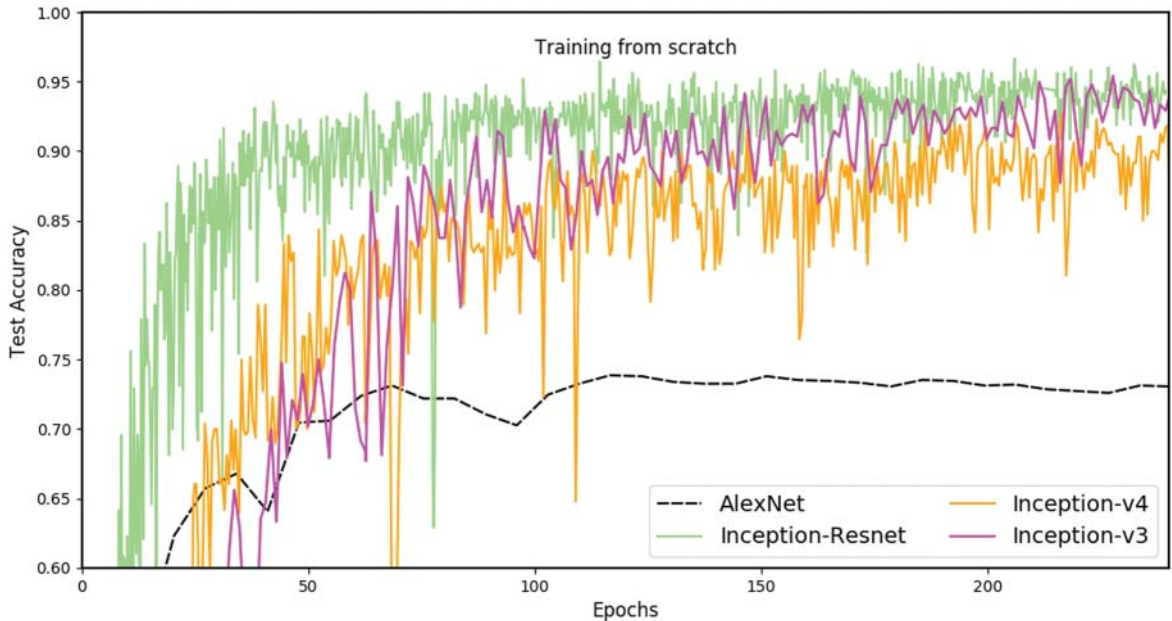
The architectures used are Inception-v3, Inception-v4, and Inception-Resnet-v2 [19, 20, 21]. The core idea under this family of networks is the *inception module*: it consists of parallel convolutional layers of different kernel sizes, which improve the ability of the network to efficiently detect features of objects having different dimensions.

All the computations shown in this work were performed on the C3E *Cloud Computing Environment* of COSILT [22], from now on called C3HPC, located in Tolmezzo (Italy) and managed by eXact Lab srl [23], equipped with two Tesla K20 Nvidia Graphics Processing Units (GPUs) loading the Nvidia CUDA Deep Neural Network library (cuDNN).

### 2.1 Training from Scratch on SEM Data Set: Comparison between Different Architectures

In the field of deep learning applied to computer vision, the first CNN that outperformed and won the *ImageNet ILSVRC 2012* was AlexNet [24]. We initially trained from scratch this simple model on our SEM data set and reached 73% of accuracy. This result was not impressive; thus, as a next step, we trained the version v3, v4, and Resnet-v2 (hereafter Resnet) of the more recent Inception architecture [19, 20] on the SEM data set.

In Figure 1, the accuracy computed on the test set up to 240 training epochs is shown for the mentioned architectures. As expected, the Inception family of networks reaches a remarkably higher value than the simpler AlexNet. It seems that Inception-v4 performs only slightly worse than Inception-v3 and Inception-Resnet on the SEM data set. However, 240 epochs were not enough for Inception-v4 to converge, as the loss function had not reached a stable minimal value yet. Moreover, it required more than twice the time needed by the other networks to reach the same accuracy in a stable way: ~ 160 hours, with respect to ~ 70 hours. After these considerations, we decided to rule out the training from scratch of Inception-v4.



**Figure 1.** Test accuracy as a function of the number of training epochs obtained by training from scratch Inception-v3 (magenta), Inception-v4 (orange), Inception-Resnet (green), and AlexNet (black) on SEM data set. All the models were trained with the best combination of hyperparameters, according to the memory capability of the available hardware.

## 2.2 Feature Extraction and Fine Tuning from ImageNet to SEM Data Set

Transfer learning is becoming a very popular technique in deep learning. It is based on the idea of storing the knowledge learned from one task and applying it to a different but related one [25]. This approach is faster than training from scratch, but the results might be less accurate in some cases (e.g., feature extraction), and the architectures which can be used are restricted to the pre-trained checkpoints available in the literature.

In this work, transfer learning was tested on our target SEM data set, using Inception-v3, Inception-v4, and Inception-ResNet checkpoints pre-trained either on ImageNet ILSVRC 2012 [17] or on SEM data set (Section 2.1).

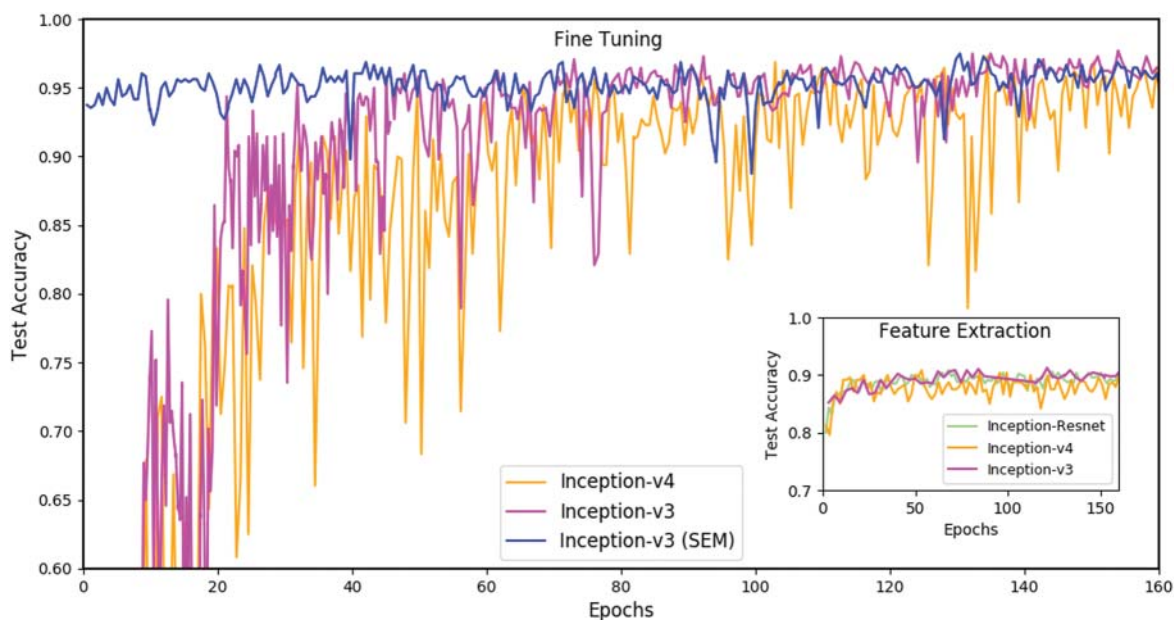
As shown in the inset of Figure 2, feature extraction accuracy does not exceed 90% for any CNN architecture. This reproduces the results obtained by [13], and confirms the limits of this transfer learning method.

Fine tuning reveals itself as a more successful technique, increasing the test accuracy to  $\sim 97\%$  as can be seen in the main panel of Figure 2 (magenta and orange lines). Due to its complexity (in terms of floating-point operations), Inception-Resnet architecture gave rise to GPU memory issues when setting the batch

size  $bs > 16$ . This made the training slower with respect to the other architectures, because of the greater number of back-propagations performed. Thus, we decided to address the complete investigation of Inception-Resnet in the future on different hardware equipment, and to omit the results in order to avoid confusion.

Having excluded the training from scratch of Inception-v4 and the fine tuning of Inception-Resnet, we finally fine tuned Inception-v3 using the SEM checkpoint obtained in Section 2.1 by training from scratch this architecture on the SEM data set. The test accuracy, shown in Figure 2 (blue line), is comparable with the ones obtained from the ImageNet checkpoint. These results lead us to the conclusion that the SEM data set is complete enough to allow the autonomous training of deep neural networks, without the need to only rely on the models pre-trained on huge data sets such as ImageNet.

Thus, Inception-v3 fine tuned on the SEM data set revealed to be the most suited architecture for our purposes; all the analyses presented in the rest of the paper have been done using this model.



**Figure 2.** Main: Test accuracy as a function of the number of training epochs obtained when fine tuning on the SEM data set Inception-v3 (magenta) and Inception-v4 (orange) starting from the ImageNet checkpoint, and Inception-v3 (blue) from the SEM checkpoint that, as expected, converges very rapidly. Inset: Test accuracy as a function of the number of training epochs obtained when performing feature extraction of Inception-v3 (magenta), Inception-v4 (orange), and Inception-Resnet (green) on the SEM data set starting from the ImageNet checkpoint. All the models were trained with the best combination of hyperparameters, according to the memory capability of the hardware available.

### 3. UNSUPERVISED LEARNING

In this section we report on our first attempt to use unsupervised techniques to automatically identify new categories within our data set [16]. The data set can be seen as a cloud of points in a space of dimension given by the number of total pixels (namely  $1024 \times 768$  for the vast majority of the images). Pattern detection within data (i.e., to detect common unspecified features in the images) in such a high dimensional space is a challenging task due to the sparsity of data in the input space. This is generally referred to as the *curse of dimensionality*.

Our approach to deal with this issue comprises three different steps. We first select a subset of SEM images captured at the same resolution to facilitate detection of hidden structures at the same level of magnification. We then apply a procedure to reduce the dimensionality of this subset. Finally, we adopt several clustering methods and quantitatively evaluate the different algorithms. In the following subsections, we describe in detail each step and discuss the results obtained so far.

#### 3.1 The $1\mu\text{--}2\mu$ Data Set Selection

The SEM data set includes images of objects whose size can vary over several orders of magnitude, ranging from 1 mm to 1 nm. Even the same objects, imaged at different levels of magnification, may exhibit different high-level features. For this reason, the image resolution is a fundamental quantity to split the data set by scale. This piece of information is not available as metadata for all the images; nevertheless, it can be recovered as it is annotated on each image on a stripe containing different information. In order to read and store the scale data, an algorithm was implemented using the OCR engine Tesseract v3.04.01 and the library OpenCV v3.4.1 for image segmentation and contours detection. Using the above algorithm (as outlined in [16]) we classify all the images in different bins of scale. In particular, the bins of  $1\mu\text{m}$  and  $2\mu\text{m}$  count a substantial amount of images and a non-zero population on all the original 10 categories. For this reason, we select the images in the interval from  $1\mu\text{m}$  to  $2\mu\text{m}$  to form a data set, which we will refer to as  $1\mu\text{--}2\mu$  data set. All the analyses described in the rest of the paper have been performed on it. Within the 52,682 images in this data set, 7,557 appear in the SEM data set and so have a hand-assigned label. Table 1 shows the breakdown according to the 10 categories.

**Table 1.** Number of images for each label in the  $1\mu\text{--}2\mu$  data set, adopting the same labelling used in [11, 12, 13], reported here for completeness: 0 = Porous sponges, 1 = Patterned surfaces, 2 = Particles, 3 = Films and coated surfaces, 4 = Powders, 5 = Tips, 6 = Nanowires, 7 = Biological, 8 = MEMS devices and electrodes, 9 = Fibres.

Label	0	1	2	3	4	5	6	7	8	9	TOT
Images	55	1,403	933	147	236	713	1,891	588	1,572	19	7,557



### 3.2 Feature Learning

To perform clusters analysis on the data set, a notion of distance between images is needed. In the literature there are several definitions of similarity between images (see for example [26, 27]); however, most of them do not perform content analysis, precluding the possibility of detecting similarities among different objects. A different way to proceed is to pre-process the images by selecting the high-level features that most characterize their content. The advantages of this approach are twofold. It highlights the most meaningful features of the images, and it helps to bypass the curse of dimensionality.

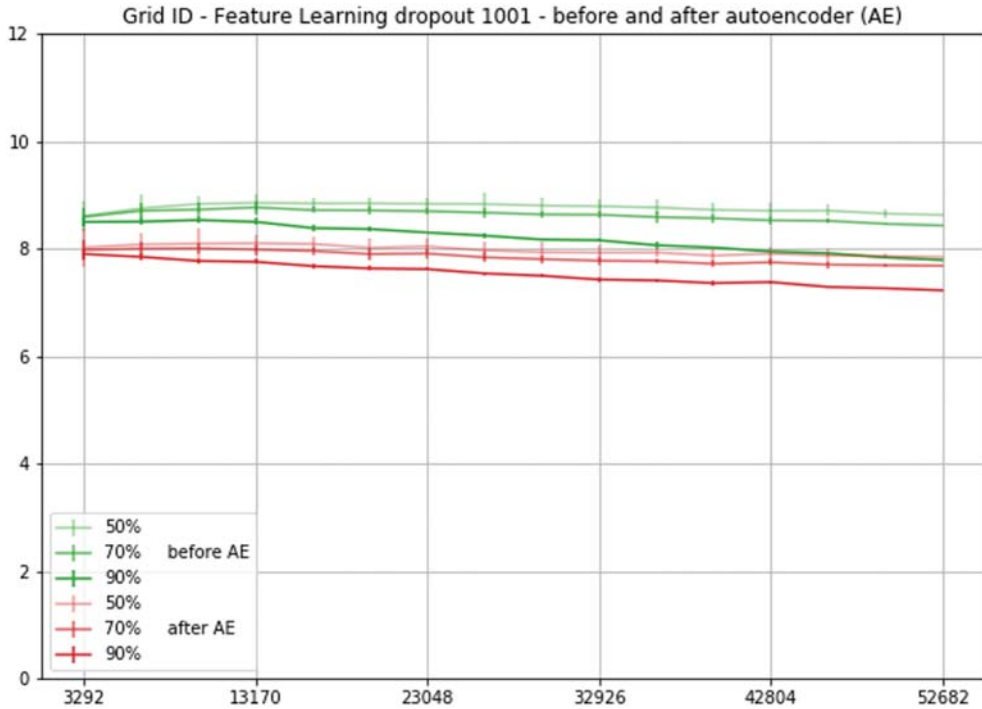
We therefore design an intermediate procedure between supervised and unsupervised learning methods. The key idea is to exploit the best performing architecture described in Section 2.2 (i.e., Inception-v3) to extract from the images in the  $1\mu\text{--}2\mu$  data set the features obtained in the last fully connected layer. For sake of clarity, we do not discuss here the different choices of model-layer that can be performed and we focus exclusively on the last layer, namely, the Logits layer, of the model fine-tuned on the SEM data set (refer to [16] for an analysis of different cases). Because of the original design of the network, suited for the prediction of the 1,001 classes of the ImageNet data set, the output of this layer has dimension 1,001. This is why the data set obtained by this procedure is referred to as the  $1\mu\text{--}2\mu_{1001}$  data set.

### 3.3 Intrinsic Dimension and Dimensional Reduction

The 1,001 features considered in the  $1\mu\text{--}2\mu_{1001}$  data set are still too many to bypass the curse of dimensionality. To further shrink the features space without losing essential information, we perform a nonlinear dimensional reduction using autoencoders [28, 29]. Autoencoders are a class of neural networks which attempt to compress the information of the input variables into a reduced dimensional space, the so-called *coding space*, and then recreate the input data set.

It should be recognized that the features considered in  $1\mu\text{--}2\mu_{1001}$  were extracted by a network trained to distinguish the 10 classes in the SEM data set. Thus, in the specific case we are analyzing, we could already assume 10 to be a reasonable coding dimension. However, in a more general framework (models trained on different data sets or features coming from lower layers), the coding dimension is a parameter that has to be set carefully. A good way to proceed is to estimate the Intrinsic Dimension (ID) of the data set via the so-called 2-Nearest Neighbors (2-NN) algorithm recently presented in [30].

The green lines in Figure 3 summarize the results obtained by 2-NN estimator on the  $1\mu\text{--}2\mu_{1001}$  data set. The Figure confirms our assumption: the lines show an evident plateau around 9, i.e., quite close to the number of the original categories. Moreover, applying an autoencoder with coding dimension 10 does not have a great impact on the ID of the data, as shown by the red lines. To maintain consistency in the naming system, we refer to the reduced representations of the data set obtained by the autoencoder as  $1\mu\text{--}2\mu_{1001\_10}$ .



**Figure 3.** Intrinsic Dimension of the 1 $\mu$ -2 $\mu$ \_1001 data set, varying the sample size, computed before autoencoding (green lines) and after autoencoding (red lines). The three brightness levels for each color correspond to the percentage of points used in the linear fit: 90%, 70%, and 50%.

We finally evaluate whether the dimensional reduction we performed on the 1 $\mu$ -2 $\mu$ \_1001 data set still provides meaningful information with respect to the original 10 categories. In order to do so, we sample 400 images from it and compare the Euclidean distances induced by their reduced representations (at 1,001 and 10 features respectively), against the following discrete distance:

$$d_{disc}(x_{i'}, x_i) = \delta_{i', i_i} \quad (1)$$

where  $i_i$  is the label of the image  $x_i$ .

Figure 4 shows the heatmap of the discrete distance on the 400 sampled images sorted by category. The black squared blocks on the diagonal represent the zero distance pairs within the same category.



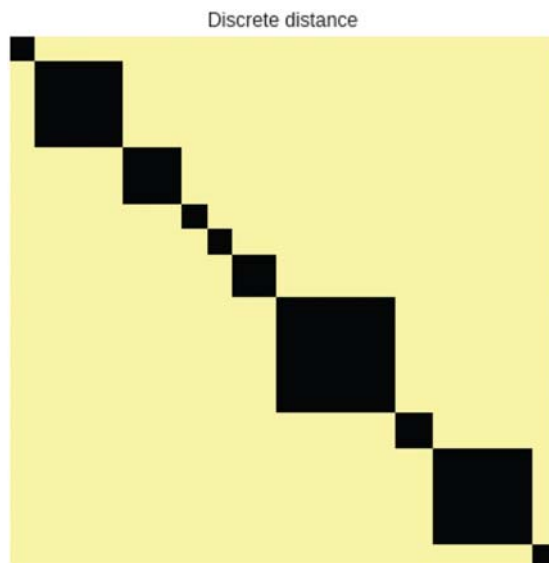


Figure 4.  $d_{disc}$  heatmap for a manually labelled subset of images.

The heatmaps of the the Euclidean distances inherited by the reduced representations before (left panel) and after (right panel) applying the autoencoder are displayed in Figure 5. In both cases, the block structure along the diagonal emerges clearly, being more definite on the right panel. This is confirmed by a slightly larger correlation index between  $d_{disc}$  and the distance after autoencoders.

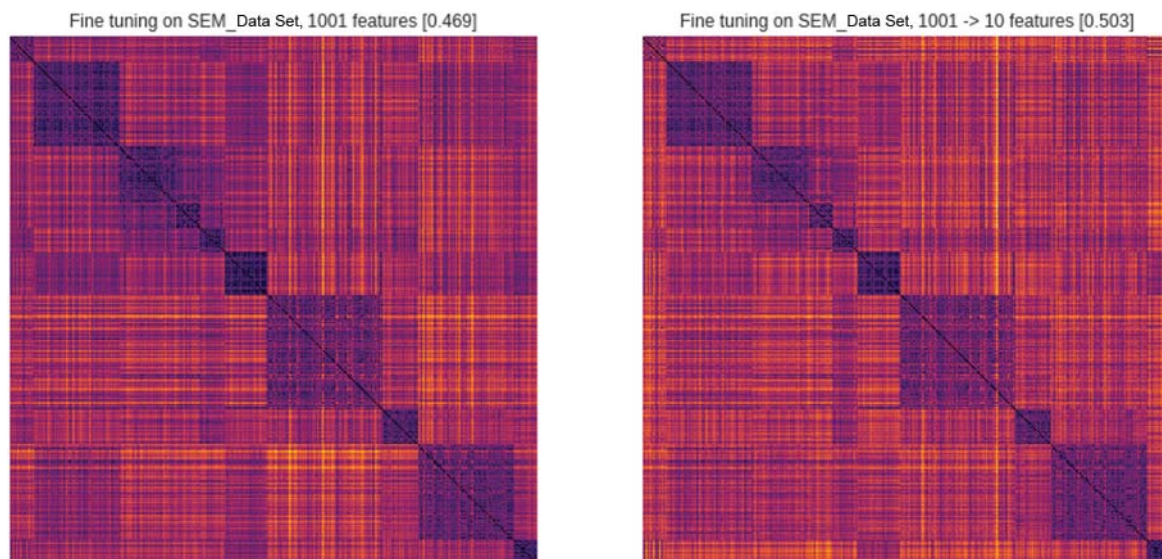


Figure 5. Heatmaps of the distances obtained via Inception-v3. The image captions specify the methods used and indicate the correlation index with  $d_{disc}$ .

### 3.4 Clustering Analysis

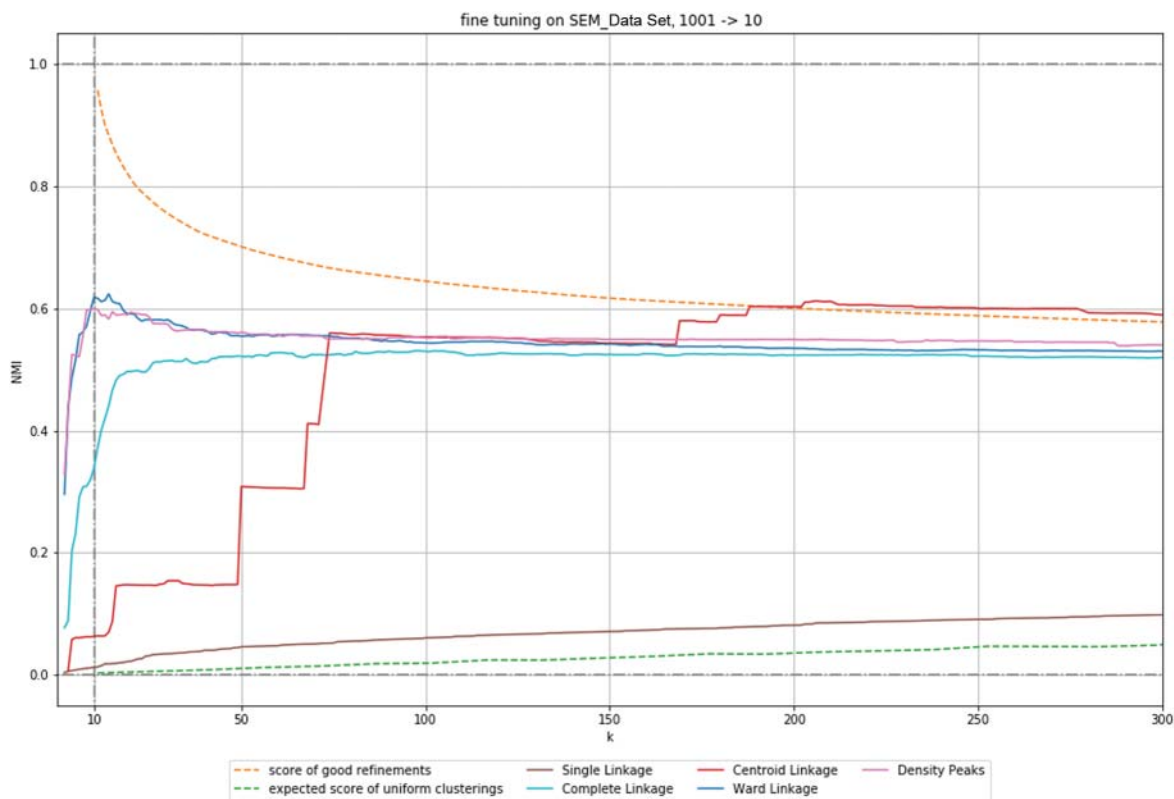
The clustering analysis on the  $1\mu\text{--}2\mu\text{--}1001\text{--}10$  data set is possible with a moderate computational effort. Among the several different algorithms available in the literature, we focus on the hierarchical agglomerative clustering methods defined by four classic linkages criteria: single, complete, centroid and Ward. Moreover, we also completed the analysis including a hierarchical version of a recently introduced density-based technique called density peaks [31].

To evaluate the quality of the clusters obtained at a given level of a hierarchy, we compare them to the original 10 categories of the SEM data set [11] via the widely adopted Normalized Mutual Information (NMI) [32]. As already specified in Section 3.1, not all images in  $1\mu\text{--}2\mu$  were classified, and thus our evaluation takes into account only the 7,557 images in the data set that comes with a hand-assigned label.

To better understand the scores of a hierarchical algorithm at different levels of the hierarchy we compare them to the scores obtained by two artificial scenarios. The first one, called *good scenario*, is constructed recursively in a divisive way: starting from the partition of the 7,557 images provided by the 10 categories, at each step the biggest cluster is split evenly in two clusters. On the other hand, the *uniform scenario* is obtained by a uniform assignment of  $k$  labels, for  $10 < k \leq 7557$ . We can then compute the NMI for both scenarios and plot them as a function of the number of clusters. We remark that those are not bounds for the NMI scores, but should be considered as a reference to help us evaluate the scores of the clustering method adopted in this study.

Figure 6 shows the NMI scores (on the labelled data only) of the clustering obtained by the five hierarchical algorithms applied on the whole  $1\mu\text{--}2\mu\text{--}1001\text{--}10$  data set. We also report the scores of the artificial scenarios discussed above as dashed lines: the orange one refers to the good case, while the green one represents the uniform case.

From Figure 6 we can spot some interesting patterns. The Single Linkage (brown line) provides the worst results and this can be explained looking at the clusters' cardinalities. This algorithm produces a big cluster merging small clusters or singletons into it. On the other hand, the Complete and Ward linkages (cyan and blue line, respectively) as well as Density Peaks (pink line) behave similarly and present good results. Actually, the latter two produce almost identical scores, and a peak is observed around 10. This behavior is due to the strong bias towards the 10 categories inherited by the reduced representation from the model used to extract the high-level features. Nevertheless, the most interesting and impressive results are the ones obtained by the Centroid Linkage (red line). Although poor scores are returned by this algorithm for a small number of clusters, they rapidly grow after  $k \sim 70$  and they outperform the results obtained by the artificial refinements used as a reference of good scores.



**Figure 6.** NMI scores of the clustering obtained by the five hierarchical algorithms (solid lines) considered as a function of  $k$ , the number of clusters. The scores of the artificial scenarios are reported as orange (good case) and green (uniform case) dashed lines.

The results summarized in Figure 6 are not only a mere evaluation of the adopted clustering algorithm, but give interesting indications to further exploit the data set. The most remarkable hint is the impressive results obtained by the Centroid Linkage (red line) around 200 clusters; this result encourages us to improve the procedure in order to classify the SEM images in a tree structure of sub-categories with small or even any manual labeling effort. The 200 clusters identified have to be scientifically validated, i.e., it must be checked whether the features characterizing the clusters have a scientific, useful meaning for SEM users; this evaluation is currently in progress at CNR-IOM in Trieste.

#### 4. CONCLUSIONS

In this paper, we presented an overview of our recent work on the classification of images at the nanoscale collected using the SEM.

In Section 2 we applied state-of-the-art deep learning techniques and CNNs on a recently published data set composed by SEM images. We performed a comparison of different Inception architectures and learning methods: training from scratch, feature extraction, and fine tuning. Feature extraction and fine tuning can be even combined together. First, feature extraction can be applied on the target data set from a pre-trained checkpoint, and then the entire model can be fine tuned. In this case, the fine tuning starts with the last layer weights initialized to the values obtained from the previous training phase and it should take less time to converge. On the other side, feature extraction may be applied after fine tuning to refine the training weights and to improve (by a small fraction) the final accuracy. However, when the number of classes is not huge, as in our case (where 10 classes are under consideration), we verified that there are no relevant gains in performance. We showed the Inception-v3 architecture trained from scratch on the SEM data set and then fine tuned to be the best choice in our case, targeting both accuracy and time to solution: it outperformed the other architectures and even the previous work on the same data set [13].

In Section 3 we presented a possible strategy to detect intrinsic structures in the SEM images in a semi-supervised way. The approach was defined semi-supervised because it was focused on the 10 features corresponding to the categories of the SEM data set, previously learned. We performed a dimensional reduction on a subset of images at similar scale by means of autoencoders, to be able to apply several clustering algorithms. NMI coefficient was then used to score them, keeping the classification in 10 categories as ground truth. The scores reached by the Centroid Linkage showed a huge evidence of a potential refinement of the current classification.

These encouraging preliminary results are the starting point for several actions, which are already ongoing: we are evaluating more in detail the dimensional reduction algorithm, to devise a general strategy to pass from the features extracted by supervised learning to a clustering method.

Moreover, we are performing the procedure described above (feature learning followed by clustering) on single categories. In this case, the number of labelled images could be increased by including the predictions realized by the CNN in [13] (even though this model achieves a worse overall accuracy than the one used in this article, its confusion matrix reports a better accuracy on the less represented categories). This will provide a tool for the automatic classification of subcategories, which we will be able to compare with the results obtained manually.

## ACKNOWLEDGEMENTS

This work has been done within the NFFA-EUROPE project and has received funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement No. 654360 NFFA-EUROPE. The authors thank A. Cazzaniga for his contribution in preparing the final version of the plots.

## AUTHOR CONTRIBUTIONS

S. Cozzini (cozzini@areasciencepark.it) and R. Aversa (rossella.aversa@kit.edu) planned the overall work. C. De Nobili (denobili16@gmail.com) and R. Aversa performed the supervised approach. P. Coronica (pc620@cam.ac.uk) performed the unsupervised approach. R. Aversa wrote the manuscript. S. Cozzini supervised the whole work. All the authors reviewed the manuscript.

## REFERENCES

- [1] G. Roughton, A.S. Varde, S. Robila, & J. Liang. A feature-based approach for processing nanoscale images. In: *Proceeding of Scanning Microscopy*, 2010, pp. 1–9. doi: 10.1117/12.853412.
- [2] A. Amani & D. Mohammadyani. Artificial neural networks: Applications in nanotechnology. In: C.L.P. Hui (edited) *Artificial Neural Networks—Application*. Rijeka, Croatia: InTech, 2011, pp. 465–478.
- [3] M.P. Nikiforov, V.V. Reukov, G.L. Thompson, A.A. Vertege, S. Guo, S.V. Kalinin, & S. Jesse. Functional recognition imaging using artificial neural networks: Applications to rapid cellular identification via broadband electromechanical response. *Nanotechnology* 20 (2009), No. 40.
- [4] M.A. Al-Khedher, C. Pezeshki, J.L. McHale, & F.J. Knorr. Quality classification via Raman identification and SEM analysis of carbon nanotube bundles using artificial neural networks. *Nanotechnology* 18(2007), No. 35.
- [5] W. Xie, J.A. Noble, & A. Zisserman. Microscopy cell counting with fully convolutional regression networks. In: *Deep Learning Workshop in MICCAI*, 2015, pp. 1–8.
- [6] H. Chen, Q. Dou, X. Wang, J. Qin, & P.A. Heng. Mitosis detection in breast cancer histology images via deep cascaded networks. In: *Proceedings of the 30th Conference on Artificial Intelligence (AAAI)*, 2016, pp. 1160–1166.
- [7] M. Mabaso, D. Withey, & B. Twala. Spot detection in microscopy images using convolutional neural network with sliding-window approach. In: *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018) 2*, 2018, pp. 67–74. doi: 10.5220/0006724200670074.
- [8] NFFA-EUROPE homepage. Available at: <http://www.nffa.eu/about/>.
- [9] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, ... & B. Mons. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3(2016), Article No.160018. doi: 10.1038/sdata.2016.18.
- [10] CNR-IOM homepage. Available at: [www.iom.cnr.it](http://www.iom.cnr.it).
- [11] R. Aversa, M.H. Modarres, S. Cozzini, R. Ciancio, & A. Chiusole. The first annotated set of scanning electron microscopy images for nanoscience. *Scientific Data* 5 (2018), Article No. 180172. doi: 10.1038/sdata.2018.172.
- [12] R. Aversa. Scientific image processing within the NFFA-EUROPE data repository. Master thesis, International School for Advanced Studies (SISSA-ISAS), 2016. Available at: <https://iris.sissa.it/handle/20.500.11767/36168#.XOZO3y2B1Bw>.
- [13] M.H. Modarres, R. Aversa, S. Cozzini, R. Ciancio, A. Leto, & G.P. Brandino. Neural network for nanoscience scanning electron microscope image recognition. *Scientific Reports* 7 (2017), Article No. 13282. doi: 10.1038/s41598-017-13565-z.
- [14] J. Yosinski, J. Clune, Y. Bengio, & H. Lipson. How transferable are features in deep neural networks? In: *Advances in Neural Information Processing Systems (NIPS)*, 2014, pp. 1–9.

- [15] C. De Nobili. Deep learning for nanoscience scanning electron microscope image recognition. Master thesis, International School for Advanced Studies (SISSA-ISAS), 2017. Available at: <https://iris.sissa.it/handle/20.500.11767/68034#.XOZO1C2B1Bw>.
- [16] P. Coronica. Feature learning and clustering analysis for images classification. Master thesis, International School for Advanced Studies (SISSA-ISAS), 2018. Available at: <https://iris.sissa.it/handle/20.500.11767/84226#.XOZLvi2B1Bw>.
- [17] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, & L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009, 248–255. doi: 10.1109/CVPR.2009.5206848.
- [18] ImageNet Large Scale Visual Recognition Challenge (ILSVRC) homepage. Available at: <http://www.image-net.org/challenges/LSVRC>.
- [19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, ... & A. Rabinovich. Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9. doi: 10.1109/CVPR.2015.7298594.
- [20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, & Z. Wojna. Rethinking the inception architecture for computer vision. arXiv preprint. arXiv:1512.00567, 2015.
- [21] C. Szegedy, S. Ioffe, & V. Vanhoucke. Inception-v4, Inception-ResNet and the impact of residual connections on learning. In: Proceedings of the 31th AAAI Conference on Artificial Intelligence (AAAI-17), 2017, pp. 1–3.
- [22] C3HPC homepage. Available at: [www.c3hpc.it](http://www.c3hpc.it).
- [23] eXact Lab srl homepage. Available at: [www.exact-lab.it](http://www.exact-lab.it).
- [24] A. Krizhevsky, I. Sutskever, & G.E. Hinton. Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, 2002, pp. 1097–1105.
- [25] S.J. Pan & Q. Yang. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering 22(10)(2010), 1345–1359. doi: 10.1109/TKDE.2009.191.
- [26] M.P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, & M.K. Markey. Complex wavelet structural similarity: A new image similarity index. IEEE Transactions on Image Processing 18(11)(2009), 2385–2401. doi: 10.1109/TIP.2009.2025923.
- [27] L. Wang, Y. Zhang, & J. Feng. On the Euclidean distance of images. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(8)(2005), pp. 1334–1339. doi: 10.1109/TPAMI.2005.165.
- [28] L. van Der Maaten, E. Postma, & J. Van den Herik. Dimensionality reduction: A comparative review. Journal of Machining Learning Research 10(2009), 66–71.
- [29] A. Géron. Hands-on machine learning with Scikit-Learn and Tensor-Flow. Boston, MA: O'Reilly Media: , 2017. isbn: 9781491962299
- [30] E. Facco, M. d'Errico, A. Laio, & A. Rodriguez. Estimating the intrinsic dimension of datasets by a minimal neighborhood information. Scientific Report 7(1)(2017), 1–8.
- [31] A. Rodriguez & A. Laio. Clustering by fast search and find of density peaks. Science 344(6191)(2014), 1492–1496.
- [32] L. Danon, A. Diaz-Guilera, J. Duch, & A. Arenas. Comparing community structure identification. Journal of Statistical Mechanics: Theory and Experiment 2005(9)(2005), P09008.



## AUTHOR BIOGRAPHY



**Rossella Aversa** got a PhD in Astrophysics at the International School for Advanced Studies (SISSA-ISAS) in Trieste, Italy. In the same town, she finalized her studies with a Master in High Performance Computing (MHPC) and worked as postdoc at CNR-IOM for three years, acquiring experience in machine learning techniques and data management. She is currently employed at Karlsruhe Institute of Technology (KIT) in Karlsruhe, Germany.  
ORCID: 0000-0003-2534-0063



**Piero Coronica** holds a PhD in Geometry from the International School for Advanced Studies (SISSA-ISAS, Trieste) combined with further post graduate studies in High Performance Computing and data science. He currently works as part of the HPC team at the University of Cambridge's Research Computing Services assisting national and international research groups to carry out academic work. The projects he has been involved in apply state-of-the-art machine learning techniques to advance active research fields ranging from astronomy to digital humanities and medical imaging.  
ORCID: 0000-0001-8235-1899



**Cristiano De Nobili** is a theoretical particle physicist. After his PhD in statistical physics at the International School for Advanced Studies (SISSA-ISAS) (2016) and a further master in High-Performance Computing (MHPC), he has been involved in deep learning. Starting from computer vision he is now a senior deep learning scientist in the field of Natural Language Processing working on Samsung's virtual assistant. Moreover, Cristiano is a machine/deep learning instructor for several masters in both the private and academic sectors. He is an active speaker and recently gave a TEDx talk on Artificial Intelligence.  
ORCID: 0000-0002-8429-1831





**Stefano Cozzini** is presently director of the Institute of Research and Technologies at Area Science Park where he coordinates several scientific infrastructures and projects at national and international level. He has more than 20 years' experience in the area of scientific computing and HPC/Data e-infrastructures. His main scientific interests are scientific computing and machine learning techniques applied to scientific data management. He is presently actively involved in the master's degree on Data Science and Scientific Computing master at University of Trieste, Italy.

ORCID: 0000-0001-6049-5242